THE NEW ENGLAND
ASSOCIATION OF TEACHERS OF ENGLISH
ORGANIZED FEBRUARY 28, 1901
WILLIAM ALLAN NEILSON, PRESIDENT. F. W. C. HERSEY, SEC'Y AND TREAS.
CHARLES SWAIN THOMAS, EDITOR

Editorial correspondence should be sent to the Editor at Newtonville, Mass.; business correspondence should be sent to the Secretary-Treasurer at 17 Lawrence Hall, Cambridge, Mass.

LEAFLET NO. 104 | CHARLES SWAIN THOMAS, NEWTON HIGH SCHOOL | JANUARY 1913

THE HILLEGAS SCALE

At the meeting of the Association on Saturday, December 14th where the general topic was "Tests of Efficiency and Standards of Measurement in the Teaching of English Composition," the center of interest and attack was the Hillegas Scale.

As the essential characteristics of this scale have already been explained in the November *Leaflet,* I need merely repeat here that this device, worked out by Professor Milo B. Hillegas, of Teachers College, Columbia University, consists of ten selected themes varying in merit from 0 to a maximum of 937. The scale is designed to aid the corrector in affixing to any given theme under survey a value which corresponds most nearly to the value designated by the rating affixed to one of the ten Hillegas norms, the norms themselves representing the concerted judgment of many different critics. If the theme under survey falls somewhere between *Value* 585 and *Value* 675, for example, the corrector may, after due judgment, grade it 634, or, roughly speaking, 63½%.

To test the practical value of this scheme, Dr. William Setchel Learned, Joseph Lee Fellow for Research in Education at Harvard University, recently undertook a series of experiments in the Newton schools. A set of fifty papers, written by elementary, grammar, and high-school pupils, was graded subjectively by five elementary-school teachers, five grammar-school teachers, and five high-school teachers. The markers were simply asked to rate the relative value of each paper as a bit of prose composition, and to designate this subjective rating by a percentage mark ranging according to judgment, anywhere from 0 to 100%.

Three weeks later these same fifteen judges, with the Hillegas Scale before them, took these same fifty papers, and graded them in relation to the values affixed to the

compositions of the Scale. In other words, they attempted to discard their subjective estimate, and to adjust a given theme, as nearly as possible, to one of the ten Hillegas norms. They indicated the measure of the variation from this norm by the proper percentage figures. For example, if the theme seemed to be nearest in merit to No. 7 of the Scale with its affixed value of 675, but inferior to No. 7, it was graded 64%, let us say; if superior, perhaps 73%.

The interesting facts revealed by this experiment are briefly summerized by Dr. Learned:

"Marking without the scale, the judges assigned to the papers values which varied among themselves from 30% in one case to 85% in another. The average extreme variation of all fifty papers was 58%. When assigned with the scale, the ratings varied from 18% in one case to 73% in another. The average extreme variation was 44%, showing a gain in uniformity of 14%, presumably due to the scale.

The variation of the nine best judgments out of the fifteen, (i. e. the nine ratings grouped about the median value assigned to each paper.) was from 10% to 43%; their average extreme variation was 30%. Using the scale, this variation was reduced to from 7% to 32%, and the average extreme variation to 17%, showing a gain for the scale of 13%.

An analysis of the effects of the scale on the average ratings of the teachers discloses the following: Without the scale, the average ratings of the teachers for the entire fifty themes vary among themselves from 23% to 74%, or 51%; with the scale they vary from 38% to 61% or 23%, showing thus a gain, apparently due to the scale, of 28%. With the primary group, the reduction of variation in average ratings is slight—24% to 23%; with the grammar group it is greater, 39% to 23%; and with the high school group it is very marked—51½ to 13%. At the same time the average extreme variation in the ratings of the individual papers by the high school group dropped from 49% to 27%. The two closest markers of the high school group rated the papers without the scale with a difference of 9% between their average ratings. The use of the scale reduced this to 2%."

In my own opinion the scale is of little practical value, notwithstanding its revealed power to secure a nearer approach to uniformity. Even this power is less than the

deductions would at first glance indicate. These papers were marked by teachers who had met in conference and had besides freely discussed the scheme outside the formal conference. All this discussion, especially the emphasis laid upon the wide variation in judgments, had tended to place each one on his guard against minimum and maximum extremes. It is fair to assume that in the second rating a large number of both the high and the low marks would naturally have disappeared, and the wide disparity would have been eliminated without the Scale.

Nor indeed am I convinced that uniformity in judgment is always desirable. To critics in the Augustan Age most of the poetry of Browning would have been anathema. It is easily conceivable that qualities of style which one teacher would encourage another teacher would discourage, and yet this diametric view might be generally helpful to a student receiving in sequence instruction from each teacher. Certainly no faultless criterion of spiritual essence is securable by a system of averages taken at any single moment. Moreover, the Scale as it now exists, is fundamentally inadequate. Of the non-artificial samples (4 to 10) all but one—possibly two—are on subjects drawn from books, whereas the majority of our school themes are, or ought to be, on subjects drawn from life. In none of the selected types is there any reported conversation, and to adjust a composition with much conversation to any one norm in the Scale is a sheer mechanical placement rather than a satisfactory judgment.

For the same reason, it is inadequate because it attempts to measure one quality by an entirely different quality. An imaginative theme on *Musings on the Lonely Isle of Nowhere* can scarcely be satisfactorily compared to one which bears such a title as *The Latest Marconni Device,* whereas the two themes may very easily be referred to a subjective A standard. As Professor Holmes pointed out at the meeting, you cannot measure light, and warmth, and redness on the same rod. To adjust imagination, individuality, original phrasing, and subtle thought to a tangible objective norm is fundamentally impossible.

My personal attempt to use the Scale through a set of fifty papers was most disheartening. The set contained compositions ranging from the fourth grade to the last year of the high school. To adjust mere immaturity of thought to one of the illiterate norms was to err on the side of strictness; to adjust it to the high norms was to err on

the side of leniency. In a sort of fateful necessity and futile desperation I flung it somewhere toward the middle. Then, too, I felt myself being constantly harassed by two contending judgments—one urged the mark which long years of theme correcting had definitely established; the other urged a search for the Hillegas norm with its predigested value. Fifty times, therefore, I felt myself caught in suspended torture between the two poles of the magnet. Release was as easily effected through errancy as through inerrancy, and I grew careless as to the means. Woefully unscientific, I admit.

Notwithstanding all this adverse comment I nevertheless think that the work of Dr. Hillegas deserves high credit. He has emphasized the variability of existing subjective judgments, he has directed self-criticism toward our own ill-defined norms; perhaps he has even pointed out the way to something that may be sparingly applied in future practice. And for these gratuities heaped up to us, we rest his hermits.

At the editor's request, Professor Holmes, Professor Neilson, Mr. Thurber, and Dr. Learned have each written out in condensed form their opinion of the Hillegas Scale.

FROM PROFESSOR HOLMES

The need of objective standards to help us in marking compositions rests on the fact that our own subjective standards vary. When we have to do justice as strictly as may be to the pupils whose work we are rating; when we seek a sure basis for comparison of results in different schools, from different teachers, by different methods; when we wish an accurate estimate of the effectiveness of our own teaching, subjective standards fail us. We often mark for other purposes than these, and we often need nothing to supplement our own reaction; but there is plenty of use for an objective standard if we can get one.

A scale to measure "merit," undifferentiated, is of little practical value, but the tests of the Hillegas Scale in Newton show that even a general scale will have considerable effect in reducing subjective variation. They show this clearly, all defects in the method of the tests aside. We need, however, scales intrinsically better than the Hillegas Scale,—scales for special kinds of writing, scales for special qualities of style, and scales for the various school grades. Such scales will be difficult to make and rather hard to use, at least in the beginning. When to turn to a

scale and how completely to submit to it, are questions which will best be answered by teachers who know just how a scale is made, what it can do, and what they themselves are about.

It must be remembered, meanwhile, that scales are essential to the study of many educational problems, even if they prove inapplicable in the immediate work of the class-room.

FROM PROFESSOR NEILSON

A fatal defect in the Newton experiments with the Hillegas scale has been pointed out by Mr. Thurber. In the absence of specific instructions some teachers, on the first reading, applied standards of the best literature, others took the best High School work as the maximum. Such differences were bound to result in variations in rating which the application of any one scale would necessarily reduce That the Hillegas Scale reduced them does not prove it good or bad.

It is important to notice that the proper field for the application of such a scale, even when perfected, is in judging the proficiency of pupils with a view to promotion or transference from one institution to another. There are other and far better tests possible for purely teaching purposes; and it would be unfortunate if so external a method of judging results were used in class-room work, in which the teacher needs to judge his pupil's attainment with reference to more specific defects than can be revealed by any such scale. For the judgments involved in framing the Hillegas Scale were the result of a rough summation of data derived from spelling, punctuation, grammar, sentence and paragraph structure, and evidences of power of thought and imagination; that summation being made without previous agreement as to the relative importance of these various classes of data. Rough totals of this kind are valueless as a guide in teaching, though for purposes of mere classification they may help to eliminate the more eccentric ratings.

FROM MR. THURBER

Five high-school teachers in Newton, all of them experienced English teachers, corrected fifty compositions in the experiment recently conducted by Dr. Learned. Among

these teachers there is a very desided feeling that the variations in ratings assigned by them to the same papers can be accounted for by certain facts which do not appear in the statistical results. These facts ought to be clearly understood, for they strike at the heart of the whole experiment as an accurate and scientific piece of work.

In the first place, insufficient directions were given for correcting and rating the papers. Conferences to interpret just what the directions meant were prohibited. Without time or opportunity to compare notes, exchange impressions, or ask questions, these teachers were requested to mark fifty compositions according to a standard that was vague, artificial, and new. The extreme misinterpretations of this standard can well be illustrated by the fact that one teacher rated the papers according to a standard of almost literary excellence, another according to what she might fairly expect from pupils of the age and training revealed in the different compositions. In other words, one marked almost entirely objectively; the other followed her usual practice of marking subjectively. That such an extreme variation in interpreting the printed instructions was inexcusable is not now the question. As a fact, however, it largely destroys the scientific value of statistics compiled from such ratings.

In the second place, the papers themselves, ranging from the fourth grade to the senior year of the high school, admitted of the largest possible variations. Indeed, it would have been hard to collect material with more possibilities for differing estimates among teachers who for many years had corrected compositions from a much narrower field and therefore of much greater informity both in technical accuracy and in general character.

Then again, so little did the correctors understand the importance of the task assigned them that quite naturally they varied to a considerable degree in the care and time which they gave to their correcting and rating.

The second reading of the fifty compositions, now with the Hillegas Scale as a measuring standard, reduced somewhat the widely varying marks of the judges. But was it the use of the scale that produced this greater informity, or simply a little more knowledge of what was meant by the original directions? Several weeks elapsed between the two readings. During this time informal conferences were held by the teachers among themselves, in which the whole matter was threshed out. It is therefore entirely probable

that a similar approach toward uniformity would have come from a second rating without the use of the scale at all.

The high-school teachers in Newton who have experimented with the Hillegas Scale are unanimous in the opinion that it is a poor scale,—badly constructed, inadequate in scope and variety of material, unsuited to the purpose for which it was designed. They also seem to agree pretty unanimously that the scale idea as applied to the correcting of English compositions, if not actually pernicious, is impracticable. No one scale,—no twenty scales—, would be sufficient to measure even the technical form,—to say nothing of the content, the originality, the imagination, the literary charm of the infinite varieties of written work which come to every high-school teacher of English in a single month! The most baneful effect of the use of scales is that they inevitably make theme correcting more objective, and less subjective; the teacher's attention is at once focussed upon the paper and not upon the boy who wrote it,—upon abstract qualities of writing, not upon personal qualities of the writer. The Hillegas Scale, as any number of better scales, used ideally, would make it possible for any English teacher in the country to correct and mark papers exactly as well as the teacher for whom those papers were written. Such a thing, on the face of it, is absurd.

FROM DR. LEARNED

The idea of a graded scale of comparison to assist in assigning to English themes values which shall be self-explanatory and generally accepted is a new and promising suggestion. The first device for this purpose is clearly preliminary and inconclusive. It is a "blanket" scale covering everything that may be included under the term merit, and expressed, in its lower and middle terms at least, in samples which are but slightly comparable with the usual school product. Its chief virtue is the thoroughly scientific character of its construction; its chief fault is that under the most favorable conditions it still admits a legitimate variation of 25%—a minimum which swells to 50% in rating specimens to which its samples are unsuited, or when the scale is hastily or carelessly applied. That it will considerably reduce the limits of variation which appear in a purely subjective rating (i. e. the unmodified reaction of teachers to the final question: 'What is that piece of writing

worth as English prose composition?)' has been conclusively shown in the Newton tests. Compared with such ratings, its graduations offer a fairly definite estimate.

But the encouraging feature of the scale is that it invites development and improvement. A similar scale, refined to such a point as to preclude more than 10%–15% variation, and with an average effectiveness under 10%, would be of great value, and no one can reasonably affirm that that is impossible. For the purpose of record, of transfer, of examination for admission or promotion, of recommendation to employers, of conferences with parents, and as a stimulus to the pupils themselves, such a scale of quality would at once satisfy a great need. The work of the investigator who would compare school with school, method with method, is greatly handicapped by the lack of precisely this thing. Its use in the highly analytical and specialized work of the class-room is problematical, but a scale would probably make its way even here in proportion to its excellence; at least as a sort of referee, or as a measure of progress towards a concrete and visible goal.

EDITORIAL NOTES

The Association is indebted to the Massachusetts Institute of Technology for their courtesy in extending to us the free use of Huntington Hall for our December meeting.

Whatever opinion we may individually hold concerning the practical efficiency of the Hillegas Scale, we must concede to it the power of stimulating an interesting discussion—profitable discussions, too.

Dr. Long's scheme of giving two marks on a theme—A for excellence in thought and E for carelessness in matters elementary—has large possibilities for good. It is interesting to note that the dread of failure, if the E faults persisted, is the agency that eliminated the careless habits. After all, perhaps we are too lenient with misspelling. The delinquency in most cases is likely to disappear when the treatment is sufficiently drastic.

Several subscriptions to the *English Journal* have been received. Anyone who wishes*The Journal* may send his name to the Editor of the *Leaflet*. When twenty have subscribed, the names will be sent to the business manager, Mr. J. F. Hosie, Teachers College, Chicago, and he will send out the notices to the individual subscribers. The price, under this arrangement is $1.50. We need ten more.

The Association is genuinely indebted to Charles F. Richardson, Professor Emeritus of Dartmouth College, for his address,—*Is English Untaught and Unteachable?* Those of us who last Saturday surrendered to the charm of Professor Richardson's personality; will easily understand why all the old Dartmouth students insist that the question must be answered with a strong negative.

KIMBALL'S ENGLISH GRAMMAR

By Lillian G. Kimball, formerly Head of English Department, State Normal School, Oshkosh, Wis.

60 CENTS

A textbook in grammar, adapted to secondary school use, distinguished by its commonsense, in which the subject is simplified and robbed of all unnecessary and minor technicalities. For this reason it will make an immediate and convincing appeal to the pupil, as well as to the teacher. The treatment is original and interesting, while the style is simple, clear, and concise. Throughout, the practical side of the subject has received special attention, many exercises being given in which the substitution of correct forms for common errors in speech will be of great benefit in improving the pupil's language in both speaking and writing. Frequent outlines and summaries are also presented. The method of instruction is positive, calling forth the constructive attitude on the part of the pupil in practical exercises and making a continual demand upon his own initiative. The illustrative sentences have been chosen for their literary excellence as well as for their fitness for the purpose.

AMERICAN BOOK COMPANY

New York **Cincinnati** **Chicago**

Hitchcock's Rhetoric and the Study of Literature

By ALFRED M. HITCHCOCK, Head of the English Department in the Hartford Public High School

A manual for the last two years of the high school course. Part I contains a compact rhetoric based upon the terms, Purity, Clearness, Force, Beauty; a vocabulary of rhetorical and critical terms is a new feature of this section. A condensed manual of composition follows with chapters on the four forms of discourse, exposition and argument being given the prominence appropriate for junior and senior work. Part II takes up the study of masterpieces and surveys the entire field of pure literature under the headings, Reading, Literature, Poetry and Prose, Varieties of Prose, Varieties of Poetry, The Study of Prose Fiction, The Study of the Drama, The Study of the Essay, The Study of Poetry. Part III gives a condensed summary by periods of English literature,—such a survey as is called for in the new college requirements. The Appendix includes questions on a number of the more commonly studied masterpieces. As in the author's **Practice-Books,** which this manual is intended to supplement, the exercises form an important and attractive feature.

HENRY HOLT AND COMPANY

New York Boston Chicago

The World's Great Classics

As Required for Reading in Group I of the College Entrance Requirements 1915-1919

THE OLD TESTAMENT NARRATIVE
With Introduction and Notes by Alfred D. Sheffield, **Riverside Literature Series, No. 204. Cloth 75 cents, net. Postpaid.**

THE ODYSSEY
Translated into English Prose by Professor George H. Palmer of Harvard University. **Riverside Literature Series, No. 180. Cloth 75 cents, net. Postpaid.**

THE ILIAD
Translated into English Blank Verse by William Cullen Bryant. **Students' Edition. $1.00, net. Postpaid.**

THE AENEID
Translated into English Blank Verse by Theodore C. Williams, formerly Head Master of the Roxbury Latin School. **Riverside Literature Series, No. 193. Cloth 75 cents, net. Postpaid.**

HOUGHTON MIFFLIN COMPANY, Boston, New York, Chicago

The Standard English Classics

The Newest Books in This Series of 70 Volumes

Palgrave's Golden Treasury — **Stevenson's Treasure Island**
Old Testament Selections — **Lincoln, Selections**
Huxley, Selections — **David Copperfield**
Stevenson's Travels with a Donkey and An Inland Voyage

The Standard English Classics represent a distinct accomplishment in the manufacture of good books at low prices. They are unique in their appropriateness of editorial material and mechanical excellence. Their pleasing appearance and convenient size commend them at once to teacher and student. The series includes all the volumes necessary to meet the College Entrance Requirements in English and many other selections from standard English authors. New volumes are being added constantly.

Upon request the publishers will gladly send you a pamphlet containing the College Entrance Requirements in English for 1913-1919, and a complete list of the Standard English Classics.

GINN and COMPANY

29 BEACON STREET, - - BOSTON

NEW YORK CHICAGO LONDON

High School Exercises in Grammar

By Maude M. Frank, A. M., De Witt Clinton High School, New York City. 206 pp. 75 cents.

As a manual of English grammar, designed particularly for high schools, this text is intended to serve a special purpose. Its aim is to provide the material needed for the rapid, intensive work which is most practical and most profitable.

Primarily, the book is intended to be used in connection with the first or second year work in English in high schools. However, by the method of cross-references employed, the various divisions may easily be used independently in a quick, intensive review of grammar during the fourth year.

Constructive Exercises in English

By Maude M. Frank, A. M. 164 pp. 50 cents.

These exercises offer in convenient and practical form an abundance of illustrative and practice material in oral and written work that many teachers desire. Special attention is given to the development of the student's vocabulary.

Longmans, Green, & Co. Publishers

Fourth Avenue and 30th Street, . . . New York City